Toward the elimination of bias in satellite retrievals of sea surface temperature 2. Comparison with in situ measurements

Citation for published version:
Merchant, CJ & Harris, AR 1999, 'Toward the elimination of bias in satellite retrievals of sea surface temperature 2. Comparison with in situ measurements', *Journal of Geophysical Research*, vol. 104, no. C10, pp. 23579-23590. https://doi.org/10.1029/1999JC900106

Digital Object Identifier (DOI):
10.1029/1999JC900106

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Journal of Geophysical Research

JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 104, NO. C10, PAGES 23,579–23,590, OCTOBER 15, 1999

Toward the elimination of bias in satellite retrievals of sea surface temperature
2. Comparison with in situ measurements

C. J. Merchant[1]

Department of Space and Climate Physics, University College London, Mullard Space Science Laboratory, Holmbury St. Mary, Dorking, Surrey, England

A. R. Harris

U. K. Meteorological Office, Bracknell, Berks, England

Abstract. We test a new scheme for retrieving skin sea surface temperatures (SSTs) from brightness temperatures observed by the satellite-borne along-track scanning radiometer (ATSR). The main improvement is in the accuracy of SST retrievals using dual-view observations at 11 and 12 μm under conditions of high optical depth of stratospheric aerosol and high tropospheric water vapor loading. This is demonstrated by comparing 10 arc min average ATSR SSTs with in situ measurements of the Tropical Atmosphere Ocean (TAO) array of buoys between September 1991 and May 1992 (when volcanic aerosol was present in the stratosphere after the eruption of Mount Pinatubo in June 1991). The 620 coincidences of ATSR and TAO SSTs are matched within 1 hour and ~20 km. The mean and standard deviation of the difference between satellite and in situ SSTs are –0.16 K and 0.37 K, respectively, compared to –1.39 K and 0.59 K, respectively, using the prelaunch scheme. Meteorological measurements recorded with most TAO SSTs allow calculation of skin SSTs, more suitable for comparison with the "radiometric" SSTs derived from ATSR. The bias and standard deviation compared to skin SST are +0.07 K and 0.27 K, respectively. About 5% of ATSR retrievals have errors which we attribute to residual cloud contamination. The robust standard deviation describing the uncontaminated data is 0.19 K. The aerosol-robustness of the new scheme is evident in the absence of any trend in SST error corresponding to the decay of the stratospheric aerosol loading (0.00 ± 0.16 K yr^{-1}). The new scheme is shown to be unbiased by water vapor loadings between 20 and 60 kg m^{-2}. The effects of matchup and buoy error and either skin-bulk variability or skin-bulk adjustment error contribute to the standard deviations given above, suggesting that the random retrieval errors intrinsic to use of the new coefficients are smaller still.

1 Introduction

In our companion paper [*Merchant et al.*, this issue], we introduce new coefficients for retrieving sea surface temperature (SST) from observations of the first along-track scanning radiometer (ATSR), which flew on the European Space Agency's ERS-1 satellite launched in July 1991. This scheme is being used in the reprocessing during 1999 of data from the ATSR mission at the Rutherford Appleton Laboratory in England. Like the prelaunch retrieval scheme of *Závody et al.* [1995], the new coefficients are based on physical modeling of ATSR brightness temperatures under a range of sea surface conditions and atmospheric states. The reader is referred to the companion paper for details of deficiencies in the prelaunch scheme that have been addressed and for the full reference list; what follows is only a brief summary.

[1]Now at Department of Meteorology, University of Edinburgh, Edinburgh, Scotland.

Paper number 1999JC900106.
0148-0227/99/1999JC900106$09.00

The prelaunch scheme was developed assuming the stratosphere to be free of aerosol [*Závody et al.*, 1995], an assumption rendered invalid by the eruption of Mount Pinatubo (Philippines) 1 month prior to the launch of ERS-1. As a result, SST retrievals using the prelaunch scheme were biased cold [*Murray et al.*, 1998]. SSTs are obtained from a linear combination of ATSR brightness temperatures observed at two view angles and in two or three channels at different wavelengths. Two channels (11 and 12 μm) can be used for a day scene. In addition, 3.7 μm channel data are available for night scenes up to May 26, 1992, when this channel failed (3.7 μm channel data from day scenes were not transmitted, priority being given to data from a reflectance channel at 1.6 μm). We henceforth refer to "dual-2" (dual-view two-channel) and "dual-3" (dual-view three-channel) retrievals, respectively. In the prelaunch scheme the sensitivity of dual-2 retrievals to stratospheric aerosol is greater than that of dual-3 retrievals by a factor of 6 or 7. Thus the presence of Pinatubo aerosol caused a discrepancy in dual-2 and dual-3 SST retrievals of up to ~1.5 K in tropical regions during the first year of the mission. The new coefficients are designed to be insensitive to stratospheric aerosol, that is, to be "aerosol-robust." This is achieved by choosing the coefficients which multiply the brightness temperatures to be orthogonal to the changes in brightness temperature caused by stratospheric aerosol.

The retrieval coefficients for both the prelaunch and new schemes are found by regression. The brightness temperature data for the regression are calculated by radiative transfer modeling for a range of SSTs and associated atmospheric conditions. Bias in retrieved SST can arise if the calculated brightness temperatures are not representative of real observations, either because of deficiencies in the radiative transfer model used or in the range of surface and atmospheric conditions covered. Our companion paper describes how we address a number of such deficiencies in deriving the new coefficients. The most important is a change to the treatment of water vapor continuum absorption in the region of the 11 and 12 μm channels. The update makes the atmosphere (as modeled) less opaque to thermal radiation at 12 μm. Since the weight given to 12 μm channel data is greater in the dual-2 retrieval than in the dual-3 retrieval, any error in forward modeling in that waveband produces greater bias in the former than in the latter.

Consistency (lack of bias) between dual-2 and dual-3 SSTs is a strong indication of the quality of the retrievals. As we report in our companion paper, the new dual-2 and dual-3 retrievals are globally unbiased relative to each other to ~0.05 K, a considerable improvement over the corresponding figure of ~0.7 K for the prelaunch scheme. Nonetheless, this could simply mean that both dual-2 and dual-3 retrievals are equally biased with respect to truth, necessitating the validation exercise we present in this paper.

Previous ATSR validation exercises can be divided broadly into two categories. First, there are the relatively large-scale validation studies conducted by *Mutlow et al.* [1994], *Harris et al.* [1995], and *Harris and Saunders* [1996], which matched global in situ data from drifting buoys against ATSR data at 1/2° or 1 km resolution. Second, there are the more restricted but higher quality validations from specialist oceanographic cruises with radiometric measurements of oceanic skin temperature, such as *Thomas and Turner* [1995] and *Barton et al.* [1995]. Between them these yielded 30 independent validation points. Although radiometric measurement is more appropriate, the total number of such validation points available under conditions of significant stratospheric aerosol loading is too small for a convincing comparison.

2 Validation

2.1. Overview

The principal novel features of the new retrieval coefficients are that they are designed to be aerosol robust and are derived with an updated treatment of the absorption properties of water vapor. Thus the priority is to test the new retrieval coefficients under conditions of high stratospheric aerosol and water vapor loading. The highest stratospheric aerosol loadings following Pinatubo were in tropical regions, which are also regions of high water vapor loading. For this reason, we choose the bulk SST measurements of the buoys of the Tropical Atmosphere-Ocean (TAO) array [*Hayes et al.*, 1991] as the in situ data source. The ATSR data consist of all the 10 arc min averaged cloud-cleared brightness temperatures (ABTs) which include all channels and are coincident with TAO SSTs.

Two objectives exist for this validation exercise. The first is the assessment of the SST retrieval process as a whole in deriving 10 arc min averaged SST products. The second is the assessment of the intrinsic merit of the new retrieval coefficients. To make the first assessment (section 2.3), we quantify retrieval errors in terms of conventional statistical measures. Such measures include the effects of any deficiencies in screening out observations contaminated by cloud and any other source of error such as corrupted data. Such effects were not modeled in the derivation of the retrieval coefficients, and it is unsurprising if they cause a small percentage of retrievals to be poor. To make the second assessment (section 2.4), we use statistical measures which are unaffected by the presence of a small percentage of outlying points. We consider the resulting statistics as a fair description of the retrieval-buoy differences arising from matchup errors, buoy error, skin-bulk temperature difference variability and the retrieval error intrinsic to use of the new coefficients. Sections 2.3 and 2.4 compare the performance of the new coefficients with previously published coefficients, including discussion of trends in retrieval error related to aerosol and water vapor loadings. It is not appropriate to draw conclusions about the absolute level of bias in retrievals of skin (radiometric) SST by comparing them with in situ measurements of bulk SST. For this purpose we apply a bulk-to-skin adjustment to TAO SSTs where possible (section 2.5), allowing a final assessment of the absolute bias in retrieved SST.

2.2. Validation Data

TAO SSTs are measured at 1 m depth by thermistors to an instrumental accuracy of 0.03 K [*Freitag et al.*, 1994]. Advantages of the TAO data include the uniform instrumentation used and the availability with most SSTs of wind speed, relative humidity, and air temperature measurements, allowing estimation of skin-bulk differences. The TAO data used consist of hourly values for these parameters. For SST the hourly value is an average of six measurements at 10 min intervals through the previous hour. A basic assessment of the quality of the TAO SSTs was made by comparing with climatology and checking that hourly variations were reasonable. No bad TAO SST data were identified.

The ATSR data used were 10 arc min (~18 km × ~18 km at the equator) ABTs covering the period from September 9, 1991 to May 26, 1992, provided by the Rutherford Appleton Laboratory. These are generated by averaging the brightness temperatures in the pixels within each 10 arc min cell which pass various cloud tests. Observations obtained during satellite maneouvres which reduce the reliability of geolocation and colocation between forward and nadir views are excluded. TAO buoys are nominally moored at locations which lie at the corner of four ABT cells. Matchups between ABTs and TAO SSTs were identified by extracting all ABTs with corners at the relevant TAO locations and then searching the TAO files for the hourly data corresponding best to the time of the ABT observations. Thus satellite and in situ matchups consist of SSTs measured within an hour and within ~20 km. We use only nighttime data for two reasons: first, we require all ATSR channels and both views to be present for each ABT cell used to allow derivation of dual-3 SSTs; and second, this minimizes the variability of the difference between the true skin and 1 m temperatures, because of the absence of a diurnal thermocline at the time of the ERS-1 overpass (~2230 hours local time).

The total number of ABT records with matching TAO measurements is 620. Since between one and four ABTs are associated with each matched TAO record, these 620 coincidences are not fully independent. There are 218 different TAO records, of which 47, 34, 43, and 94 are associated with one, two, three and four

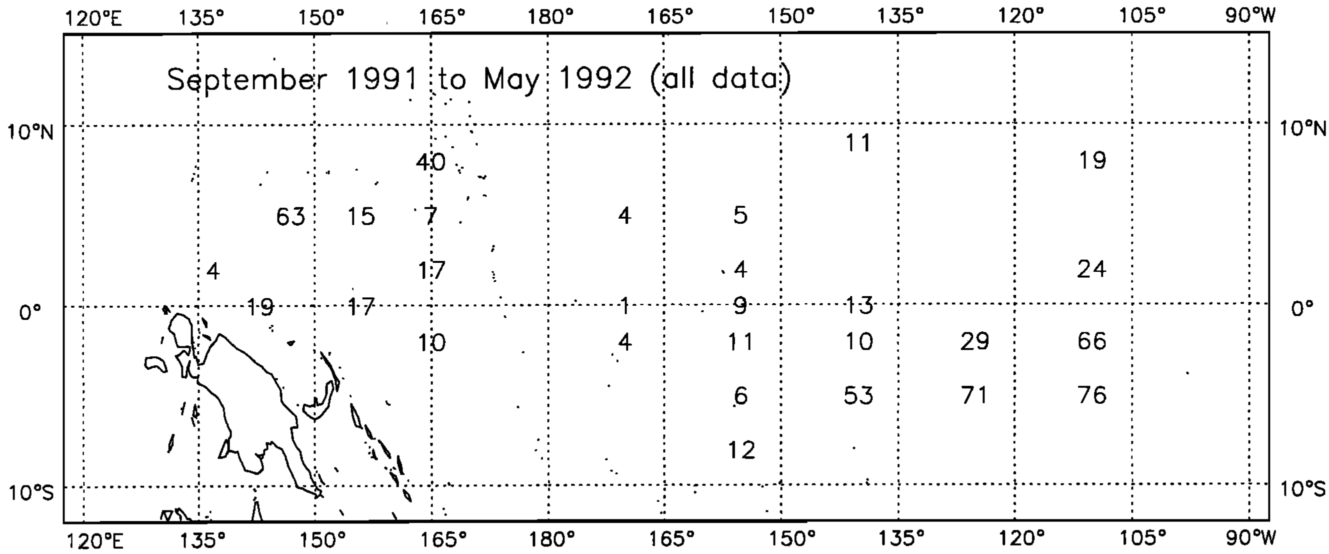


Figure 1. Number of ABTs matched with TAO records by TAO location.

ABT records, respectively.

The distribution of matched ABT records is shown in Figure 1. The TAO array of buoys evolved between September 1991 and May 1992, with the predominance of coincidences in the months to November 1991 coming from buoys clustered in the east Pacific around 120° W. Following December 1991, the gap between 140° W and 165° E was progressively filled, and buoys were added in the the west Pacific north of New Guinea, until all the locations indicated in Figure 1 had been populated by the end of May 1992.

This shifting of the buoy distribution westward toward the Warm Pool during the validation period contributes to a trend in time toward wetter atmospheres in the data set. Microwave brightness temperatures at 23.8 and 36.5 GHz from the ERS-1 microwave radiometer (ATSR/M) and associated products are available for 85% of the matched ABT records. The principal function of the ATSR/M [*Bernard et al.*, 1993] is to determine the path delay from tropospheric humidity of the radar altimeter. Estimates of water vapor column and cloud liquid water above the ground track of the satellite are generated from the microwave brightness temperatures using relationships derived by regression, with the water vapor retrieval reportedly accurate to 3 kg m^{-2} [*Eymard et al.*, 1996]. To obtain an indication of the clear-sky total water vapor loading for each ATSR-TAO coincidence, we select the nearest ATSR/M water vapor retrieval for which the corresponding cloud liquid water retrieval is low (<0.05 kg m^{-2}). All clear-sky water vapor estimates thus derived are within 300 km of the buoy location, the average separation being 129 km. These are plotted in Figure 2, where the tendency for later coincidences to occur under wetter atmospheric conditions is clear, with the apparent linear trend in the data being 35 kg m^{-2} yr^{-1}. A corresponding trend is evident if the difference between TAO SST and the 12 μm nadir brightness temperature is plotted as a proxy for total water vapor (not shown). Figure 2 also shows the wide range of water vapor loading in the validation data, from 20 to 60 kg m^{-2}. This wide range is beneficial to the testing of the updated parameterization of water vapor continuum absorption used to develop the new retrieval coefficients.

Lastly, we note that the loading of stratospheric aerosol from the Pinatubo eruption diminshes between September 1991 and May 1992. Processed data from instruments on the Upper Atmosphere Research Satellite (UARS), supplied by R. Grainger and A. Lambert of Oxford University, include the 35-day zonal mean optical depth at 12.14 μm of aerosol between 68 and 10 hPa [*Lambert et al.*, 1997]. The mean of these data between 10° N and 10° S is plotted in Figure 3. Although these data underestimate the total stratospheric aerosol optical depth (since the lower height of 68 hPa is above the tropical tropopause to avoid high equatorial cloud), they correctly indicate the magnitude of the optical depth, ~0.01, and show the reduction in optical depth over the validation period to be ~50%. We may take a value of –0.01 yr^{-1} as a rough estimate of the trend in total optical depth during the period of the validation data. Thus the extent to which the new retrieval coefficients are robust to stratospheric aerosol will be reasonably well tested.

2.3. Retrieval Process: Conventional Statistics

To place the utility of the new retrieval coefficients in context, we first present and discuss the performance of two previous schemes for the retrieval of SST from ATSR: the prelaunch scheme [*Zavody et al.*, 1995] (hereafter referred to as Z95) and the scheme of *Brown et al.* [1997] (hereafter, B97). The essential features distinguishing these are summarized in Table 1; see *Merchant et al.* [this issue] for more details. Summary statistics of the validation results for all three retrieval schemes are given in Tables 2 and 3.

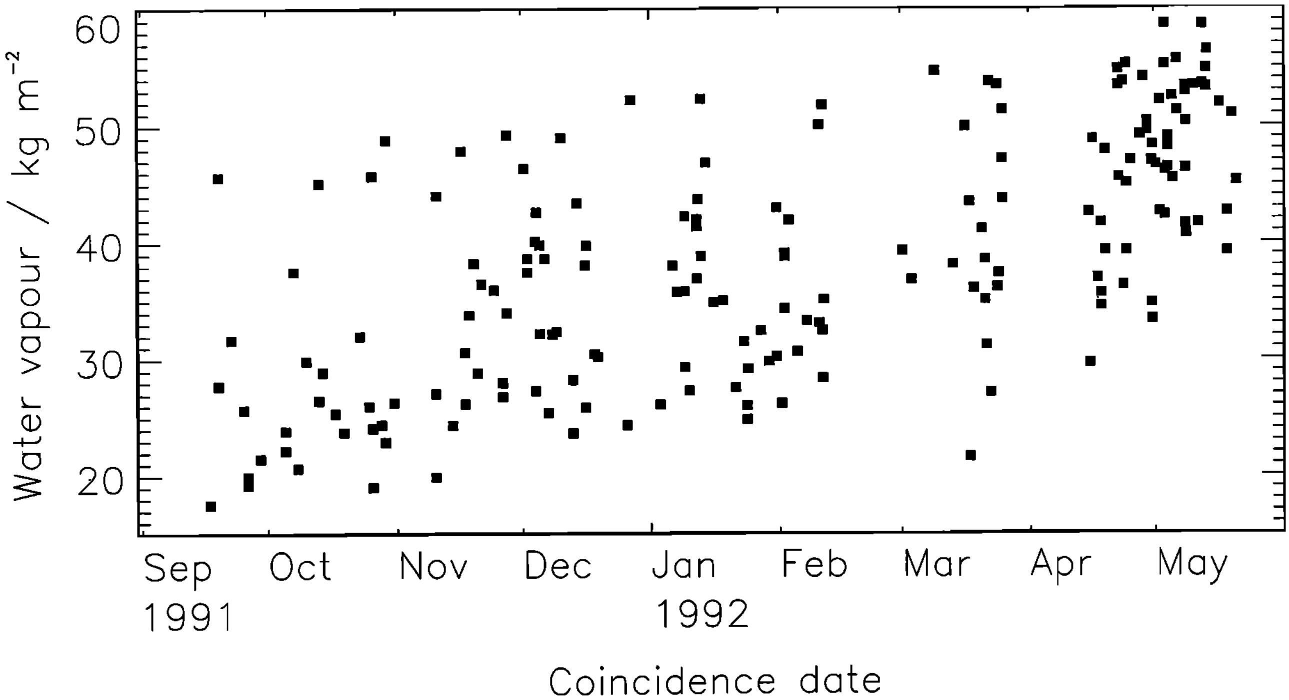


Figure 2. Clear-sky total water vapor for ATSR-TAO coincidences, estimated from microwave brightness temperatures and displaying a temporal trend.

Scatterplots of SSTs retrieved using the Z95 schemes against the in situ SSTs are shown in Figures 4a and 4b for three-channel and two-channel retrievals, respectively. (All retrievals presented here use the ATSR dual-view capability. Hereafter, three-channel and two-channel retrievals are referred to as dual-3 and dual-2 retrievals, respectively.) The line of equality is also drawn. While dual-3 retrievals are good (mean difference -0.18 K and standard deviation 0.24 K), the dual-2 retrievals suffer negative bias (mean difference -1.39 K) and larger scatter (standard deviation 0.59 K). This illustrates the advantages of the 3.7 μm channel in satellite retrieval of SST. First, the much stronger dependence of radiance on temperature at 3.7 μm than 11 or 12 μm (across the range of clear-sky brightness temperatures over ocean, ~270 to ~300 K) means that absorption of radiance by atmospheric constituents has a lesser impact on the top-of-atmosphere brightness temperature, and secondly, there is in any case less absorption by atmospheric water vapor in the 3.7 μm channel than in the 11 or 12 μm channels. However, since 3.7 μm channel data are not available for day scenes, nor for night scenes after the failure of this channel in May 1992, the biases evident in the dual-2 retrievals have adversely affected much of the ATSR SST record derived using the prelaunch coefficients.

The corresponding time series of differences between retrieved and in situ SSTs are shown in Figures 5a and 5b. We compare observed trends in the differences with those expected on the basis of the trends in aerosol optical depth and water vapor loading. Assuming the aerosol to be best described by the "aged volcanic" mode, we expect Z95 SST retrievals to warm by 0.1 K yr^{-1} (dual-3) and 0.8 K yr^{-1} (dual-2) from a trend in aerosol optical depth of -0.01 yr^{-1} (see Table 2 in our companion paper). At the same time, the 35 kg m^{-2} yr^{-1} increase in water vapor loading in the data will cool the retrievals, the expected trends being -0.1 K yr^{-1} and -0.8 K yr^{-1} for dual-3 and dual-2 retrievals, respectively (see Table 4 in our companion paper). Thus for Z95 retrievals we expect the warming trend from decreasing aerosol optical depth to be offset to some degree by the effect of the trend in water vapor, although the exact balance is uncertain because the trend in aerosol optical depth is approximate. A trend in dual-2 differences is apparent to the eye in Figure 5b. Linear regression quantifies this trend as 0.28±0.24 K yr^{-1}. (Errors on trends in this paper are quoted as ±2 standard deviations; that is, they constitute a 95% confidence interval around the estimated value. The number of degrees of freedom assumed is based on the number of ABT records; since there is dependence between points where more than one ABT record is matched to a single TAO SST, the quoted confidence intervals are likely to be biased narrow.) The trend estimate in dual-3 differences, meanwhile, is 0.09±0.10 K yr^{-1} (nonsignificant). Thus the trends in Z95 retrievals are modest, but we

Table 1. Features of Retrieval Schemes

Scheme	Stratospheric Aerosols	Water Vapor Continuum	Across-track Scheme	Latitude
Z95	not robust	original parameterization	50 km bands	different coefficients for tropical, middle, and high latitudes
B97	nearly robust	original parameterization	50 km bands	different coefficients for tropical, middle, and high latitudes
New	robust	updated parameterization	interpolation by atmospheric path length	global coefficients

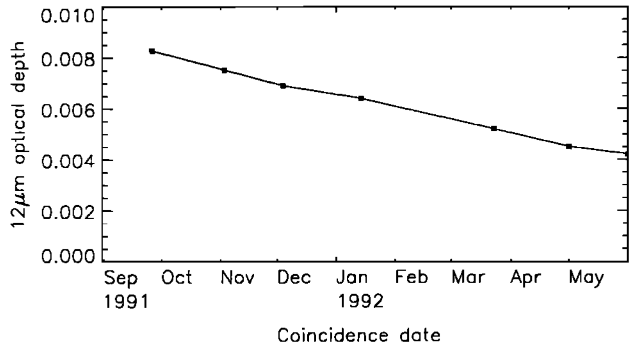


Figure 3. Mean 12.14 μm optical depth of aerosol between 68 and 10 hPa from UARS measurements.

consider that this arises partly from the effect of compensating errors.

Figures 4c and 4d are scatterplots of SSTs retrieved using dual-3 and dual-2 retrieval coefficients from B97. As before, the dual-3 retrievals are successful, with a mean difference of -0.09 K and standard deviation 0.26 K. B97 dual-2 retrievals do better than those of Z95, with smaller bias (mean difference -0.35 K) and lower scatter (standard deviation 0.41 K). This improvement comes from the robustness to aged volcanic aerosol incorporated into the B97 coefficients. In Figure 5d the time series of differences between B97 dual-2 retrievals and in situ SSTs shows the presence of a trend of -0.64 ± 0.16 K yr^{-1}. A negative trend is anticipated because B97 coefficients were derived using the same continuum parameterization as Z95, although the observed trend seems to be steeper than the -0.35 K yr^{-1} expected from this effect alone (see Table 4 of our companion paper). No significant trend is present in B97 dual-3 retrievals (0.04 ± 0.11 K yr^{-1}).

We also derive SSTs using the new dual-3 and dual-2 coefficients and compare these with the TAO SSTs in Figures 4e and 4f (scatterplots) and Figures 5e and 5f (time series of differences). As with dual-3 retrievals of earlier schemes, the new dual-3 retrievals show low bias (mean difference 0.00 K) and scatter (standard deviation 0.25 K), with no significant trend in differences over the validation period (-0.08 ± 0.10 K yr^{-1}). The new dual-2 retrievals improve on those of Z95 and B97. The mean difference is -0.16 K, the standard deviation is 0.37 K, and again there is no significant trend (-0.08 ± 0.15 K yr^{-1}). The mean difference between the dual-3 and dual-2 retrievals is 0.16 K, with dual-2 retrievals cooler. A similar feature is seen for tropical regions in the inter-algorithm comparisons presented in our companion paper (Figure 7 therein) and is a consequence of the validation being comprised of tropical data only, for the reasons mentioned in section 2.1. The shift in dual-2 retrievals of -0.08 K relative to dual-3 retrievals observed in the interalgorithm comparison between September 1991 and May 1992 is too small to be detected as a trend in differences in the validation data. The effect of using the updated water vapor continuum parameterization can be seen directly by looking at the differences against the microwave estimate of clear-sky total precipitable water (Figure 6). The B97 dual-2 SSTs are seen to be relatively cooler when the TPW is greater. This tendency is not evident in the new dual-2 SSTs, and thus the updated parameterization is clearly an improvement.

These results suggest that we have achieved our objective of developing retrieval coefficients which (1) are robust to the stratospheric aerosol from Pinatubo and (2) are not biased cold under high water vapor loadings. We cannot definitively rule out the possibility of compensating trends from sensitivity of the retrievals to the changing loadings of stratospheric aerosol and water vapor (see earlier discussion of the Z95 retrievals), although the improved ability of the new coefficients to cope with a wide range of water vapor loading (Figure 6) implies that any such sensitivity is small. While there is no demonstrated improvement in standard deviation brought by the new dual-3 coefficients over previous dual-3 coefficients, the new dual-2 coefficients represent significant progress. The statistics for the three retrieval schemes are summarized in Table 2.

2.4. Retrieval Coefficients: Robust Statistics

The statistics in Table 2 do not characterize fairly the center and spread of the distribution of differences between the retrieved and in situ SSTs because they are strongly influenced by a small percentage (5%) of outliers which do not fit the approximately Gaussian distribution of the majority of the data. Our view, based on both simulation of the retrieval process and inspection of the 1 km resolution images corresponding to the outliers, is that they arise where the ABTs observed are not true clear-sky brightness temperatures because of residual cloud contamination. Note that both cold and warm outliers can arise by this means. Residual cloud contamination usually depresses brightness temperatures in the affected view or views, that is, introduces a negative bias into some or all of the observations. The size of this negative bias differs between the ATSR channels in a manner that depends on the detailed cloud properties. Once weighted by the retrieval coefficients, which in general include both positive and negative numbers, the net bias in the retrieved SST estimate may be positive or negative. Brightness temperatures affected by residual cloud contamination were not modelled in the derivation of the retrieval coefficients, and it is unsurprising that the resulting retrievals are adversely affected (especially dual-2 retrievals). Research characterizing the non-Gaussian correlated errors potentially introduced into ABTs by such cloud contamination is ongoing.

The distributions of the differences between retrieved and in situ SSTs for retrievals made using the new coefficients are given in Figure 7. Note that one extreme outlier is omitted from the dual-2 distribution for clarity. The histograms are shown using 0.05 K bins. The dotted lines show the ideal Gaussian distribution defined by the mean and standard deviation (Table 2). In both cases, this distribution fails to capture the peak and width of the

Table 2. Conventional Statistics of Differences Between Retrieved and in situ SSTs

	Dual-3		Dual-2	
	Mean, K	S.d., K	Mean, K	S.d., K
Z95	-0.18	0.24	-1.39	0.59
B97	-0.09	0.26	-0.35	0.41
New	0.00	0.25	-0.16	0.37

Table 3. Robust Statistics of Differences Between Retrieved and in situ SSTs

	Dual-3		Dual-2	
	Mean, K	RSD, K	Mean, K	RSD, K
Z95	-0.19	0.18	-1.26	0.44
B97	-0.12	0.20	-0.28	0.28
New	-0.01	0.18	-0.14	0.20

RSD, robust standard deviation.

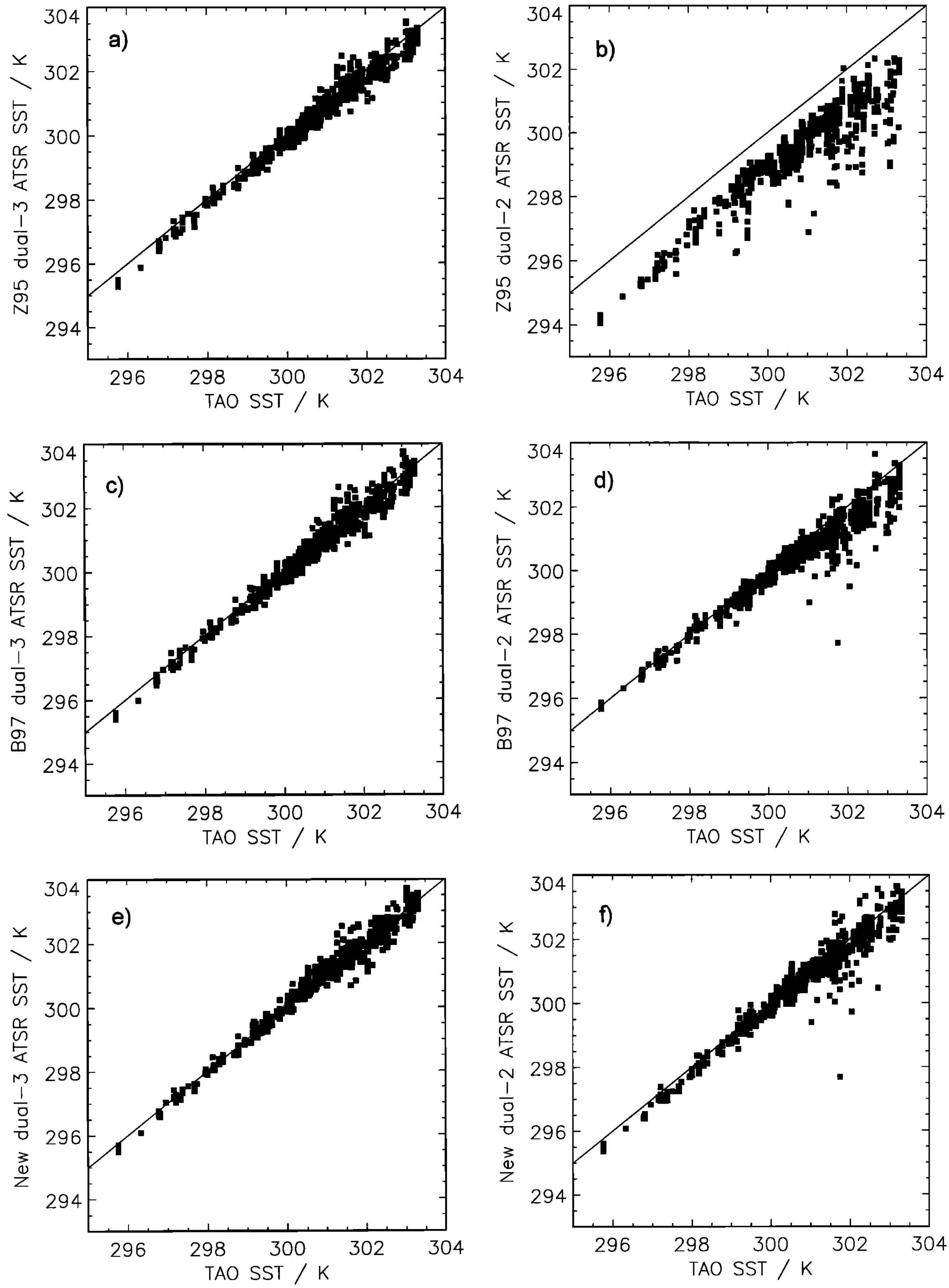


Figure 4. Comparisons of SSTs retrieved from ATSR ABTs using various retrieval schemes with matched SSTs from TAO buoys.

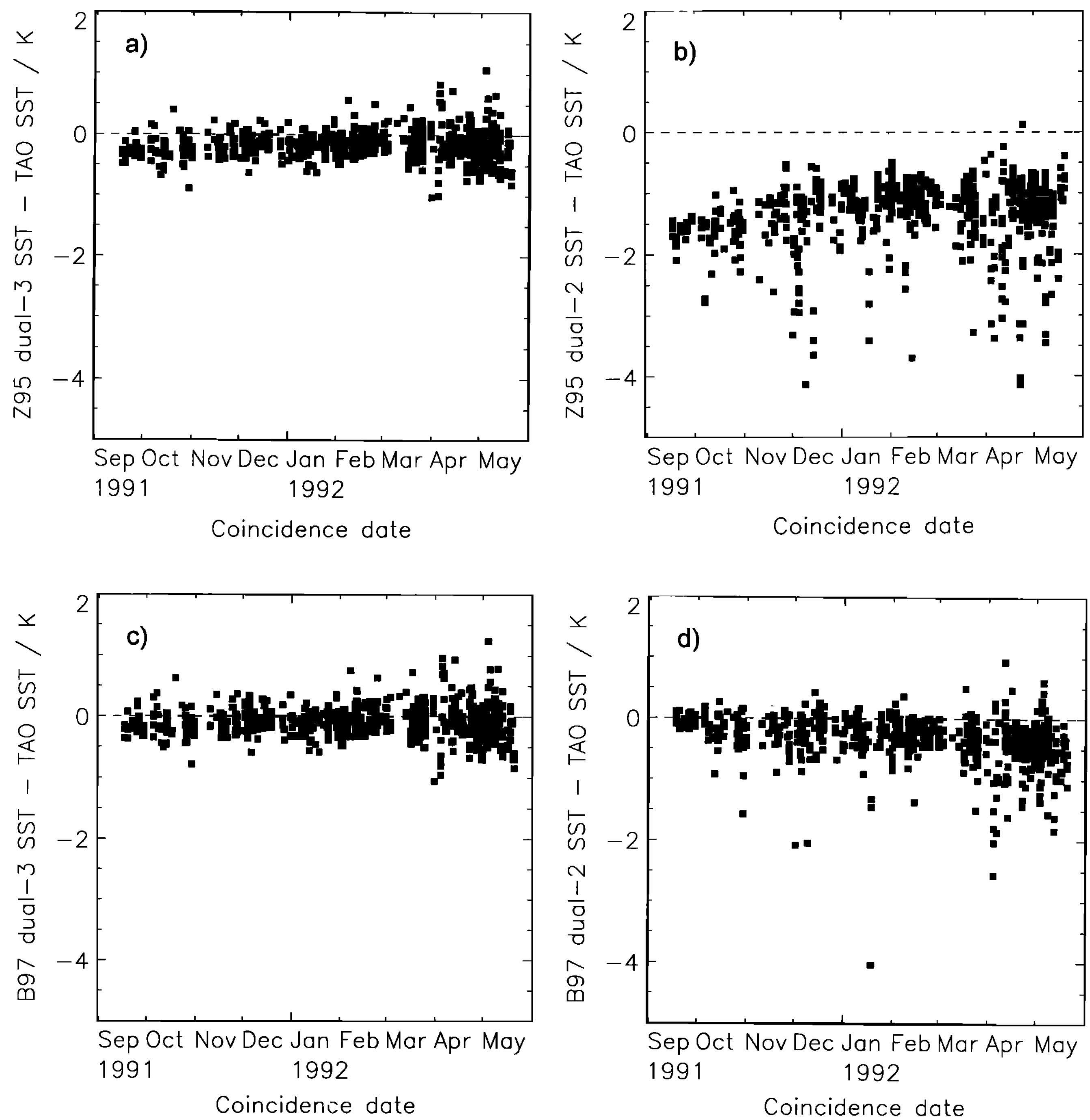


Figure 5. Time series of differences between satellite and in situ SSTs.

distribution of the bulk (~95%) of the data. The dashed lines show the Gaussian distributions defined using the median difference instead of the mean and using the range of temperature difference between the 25 and 75 percentiles to define the width. (The middle 50% of data lie between the 25 and 75 percentiles.) For a Gaussian distribution the mean and median are equal, and the 25 and 75 percentiles are 0.674 standard deviations from the mean. An estimate of the standard deviation can be obtained from data by dividing the difference of the 75 and 25 percentiles by 1.348. The exact choice of percentiles is arbitrary, within the constraints that too wide a choice makes the influence of outliers more significant, and too narrow a choice invites a large error on the estimate of standard deviation. For the distributions of residuals in this paper, choosing thresholds which span between 70% and 30% of the distribution causes fluctuations in this estimate of standard deviation of ±0.01 K, which is comparable to the error on the conventional estimates of standard deviation. The Gaussian curves obtained by this simple approach better describe the distribution of the majority of the temperature differences, as can be seen in Figure 7.

Thus, while the results in Table 2 summarize the performance of the whole retrieval process, the presence of outliers obscures the assessment of the relative merits of the various retrieval coefficients when applied as intended to true clear-sky brightness temperatures. For this latter purpose, we consider that use of statistical

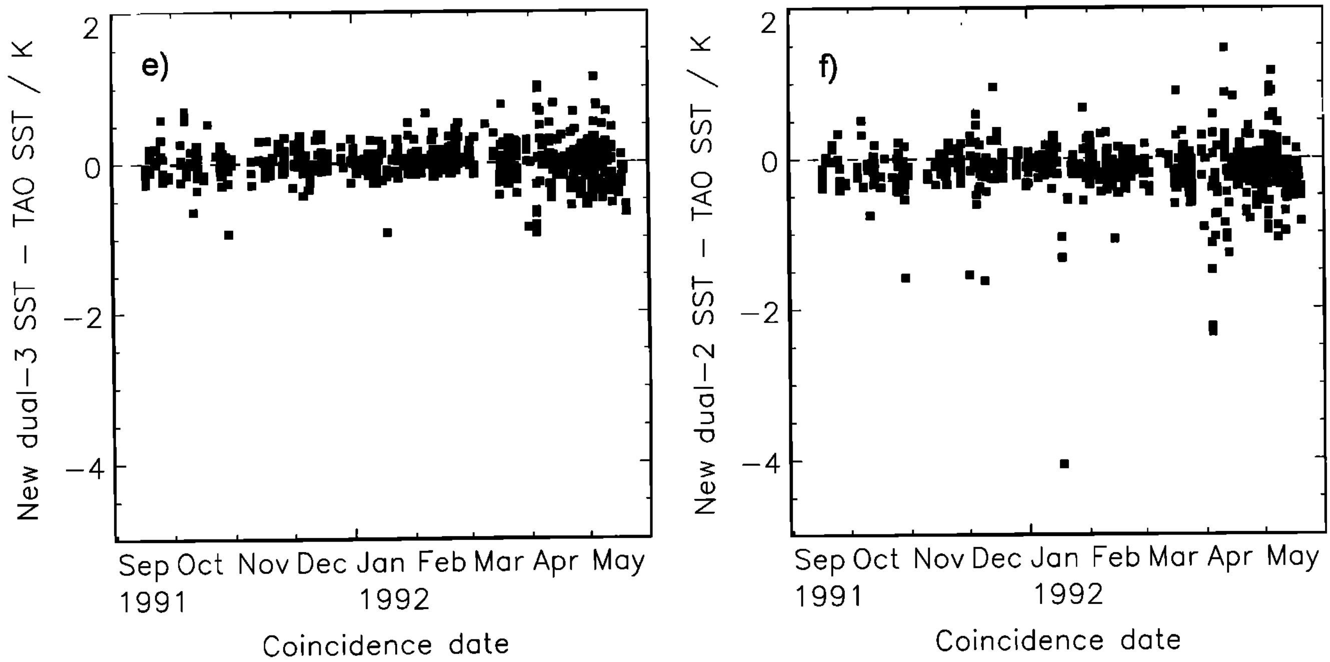


Figure 5. (continued)

measures chosen to be relatively unaffected by small numbers of outliers is more informative. Such measures are commonly referred to as being "robust" [e.g., *Analytical Methods Committee*, 1989]; the reader will be able to distinguish from the context this usage from our use of robust to indicate insensitivity of retrievals to stratospheric aerosol contamination. The median and robust standard deviation (RSD) used to define the more representative Gaussian distributions in Figure 7 are the simplest such measures. Table 3 is analogous to Table 2, but using these alternative estimators.

First, consider the results in Table 3 related to dual-3 SSTs. B97 retrievals are ~0.1 K warmer than those of Z95. We interpret this as a consequence of B97 being nearly robust to volcanic aerosol: an aerosol optical depth of 0.01 is calculated to change the Z95 SSTs by -0.11 K, and the B97 SSTs by only -0.02 K. New SSTs are warmer than B97 by a further ~0.1 K, most of which can be explained by the change in continuum parameterization and the incorporation of "full" aerosol robustness.

Turning to the results for dual-2 SSTs, Table 3 makes clear the progressive improvements in the scatter of retrievals from Z95 to B97 to this present work. The RSD reduces in going from Z95 to B97 because the latter retrievals are not as susceptible to variability in stratospheric aerosol loading. In turn, the more realistic continuum parameterization brings the RSD for retrievals with the

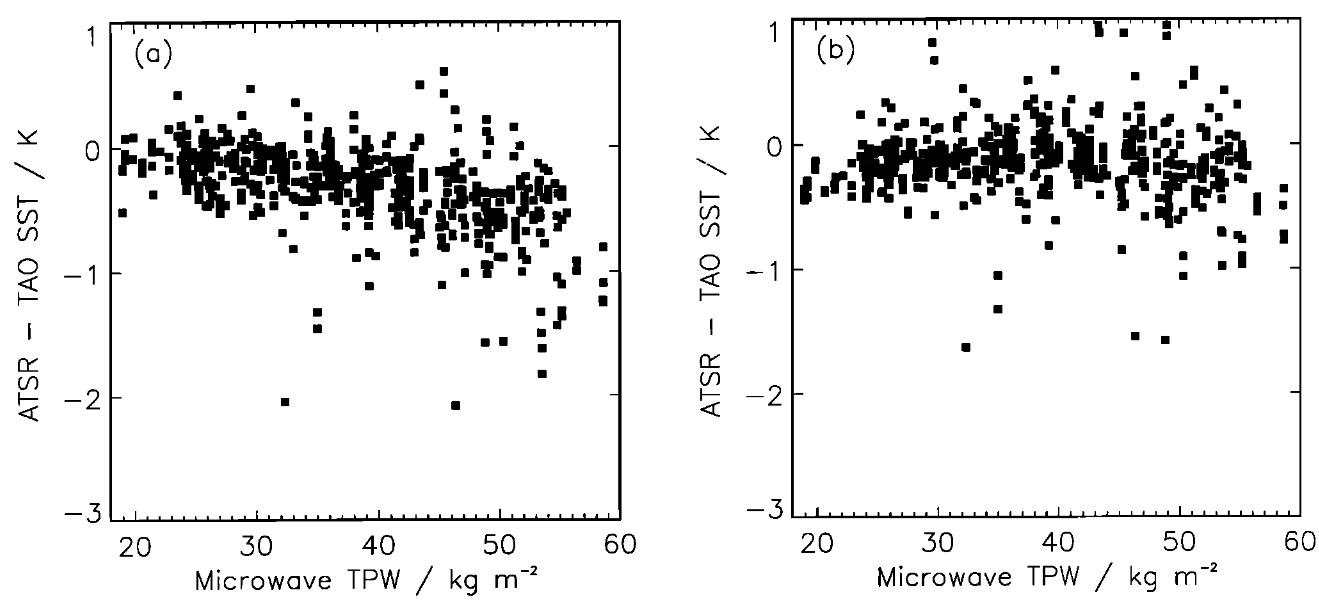
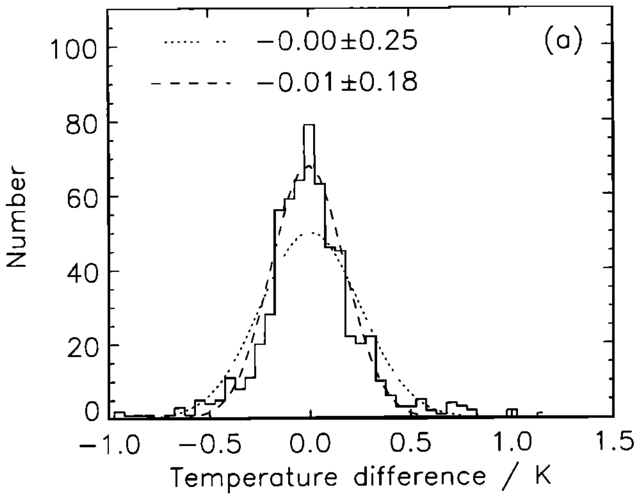


Figure 6. Differences between dual-2 retrievals and in situ SSTs against clear-sky total precipitable water. (a) ATSR SSTs using the B97 coefficients. (b) ATSR SSTs using the new coefficients, derived using an updated parameterization of water vapor continuum absorption.



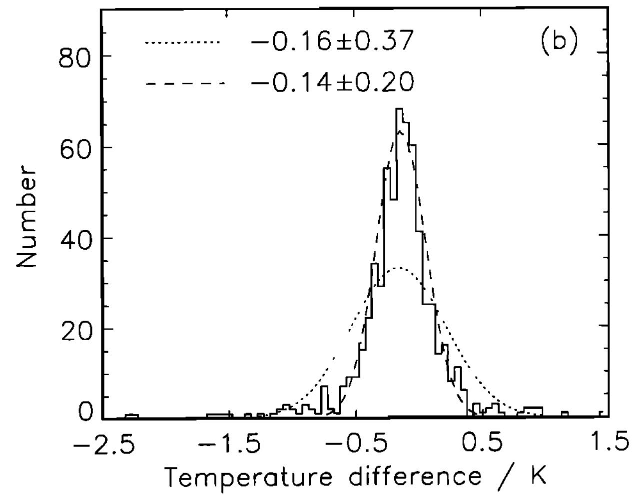


Figure 7. Distribution of differences between retrieved and in situ SSTs using (a) new dual-3 coefficients and (b) new dual-2 coefficients.

new coefficients down to 0.20 K, almost as good as the RSD for dual-3 retrievals.

2.5. Absolute Accuracy: Bulk-to-Skin Correction

Thus far we have discussed the mean and median differences between the satellite and in situ temperature measurements without drawing inferences about the absolute bias of the satellite retrievals. In order to do so, we need to account for the difference between bulk and skin SST.

The skin effect model selected for this comparison is that of *Fairall et al.* [1996a], developed for the TOGA Coupled Ocean-Atmosphere Response Experiment (COARE) region. The model is a blending of the forced- and free-convection parameterizations presented by *Saunders* [1967]. The surface fluxes of momentum, moisture and sensible heat required as inputs to the skin effect model have been calculated using the meteorological data supplied as part of the majority of the TAO data. The calculations are based on bulk parameterization scheme of *Fairall et al.* [1996b], also specifically derived for the TOGA-COARE data set. The turbulent flux scheme is based on the model described by *Liu et al.* [1979] and includes parameterizations for the cool skin effect and warming of the upper ocean by solar radiation. The latter part is not implemented in this study, as the comparisons are all performed at ~2230 hours local time, by which time convective overturning can be expected to have eroded the diurnal layer. The longwave radiation parameterization used is that of *Grant and Hignett* [1998], which again has been derived for the TOGA-COARE region and is thus expected to be reasonably representative for the data used in this study. Grant and Hignett also describe the characteristics of both their data and the effectiveness of the parameterizations of Fairall et al., in comparison with direct flux measurements and, in the latter case, those derived from eddy correlation techniques. In summary, systematic errors of the order of 10-15% are likely, and these propagate through to a systematic error in predicted skin effect of the order of 20%. Random errors arise from errors in the TAO measurements. Pessimistic estimates of error in each measurement, based on *Freitag et al.* [1994], are air temperature ~0.2 K, relative humidity ~4%, and wind speed ~1 m s^{-1}. Taking a typical air-sea temperature difference as 1 K, relative humidity of 80%, and wind speed of 8 m s^{-1}, the fractional errors in the heat and momentum fluxes are then ~25% and ~25%, respectively, propagating through to a random error in the bulk-skin adjustment ~35%.

Measurements of near-surface wind speed, air temperature and relative humidity are available for 135 ATSR-TAO coincidences, with 383 corresponding ABT records. The statistics of differences between retrieved and buoy SSTs before adjustment of the latter are shown for this subset in Table 4. This demonstrates that the statistics of the subset are similar to those of the full validation set, except that the standard deviations of differences are smaller because some extreme outliers are not present. The corresponding results after bulk-to-skin adjustment of the in situ SSTs are shown in Table 5. Note that the small reductions in s.d. after the application of bulk-skin adjustment suggests that at least some of the variability in the skin-bulk difference is correctly represented in the adjustment calculations. The scatterplots of the retrievals against adjusted SSTs are presented in Figures 8a and 8b, and the time series of residuals in Figures 8c and 8d. Note that only one of the nine TAO moorings which gave coincidences with ATSR prior to December 1991 was fully instrumented, allowing calculation of a skin SST estimate. This accounts for the scarcity of data points at the beginning of the time series (Figures 8c and 8d), which is why our analyses of trends in the various retrieval schemes are best conducted against the bulk SSTs in section 2.3.

Table 4. Statistics of Differences for Subset Before Bulk-to-Skin Adjustment

	Mean, K	S.d., K	Median, K	RSD , K	Trend, K yr^{-1}
Dual-3	0.01	0.23	-0.02	0.17	0.05±0.13
Dual-2	-0.15	0.28	-0.14	0.18	-0.08±0.16

RSD, robust standard deviation.

Table 5: Statistics of Differences for Subset After Bulk-to-Skin adjustment

	Mean, K	S.d., K	Median, K	RSD , K	Trend, K yr^{-1}
Dual-3	0.23	0.22	0.19	0.17	0.13±0.13
Dual-2	0.07	0.27	0.07	0.19	0.00±0.16

RSD, robust standard deviation.

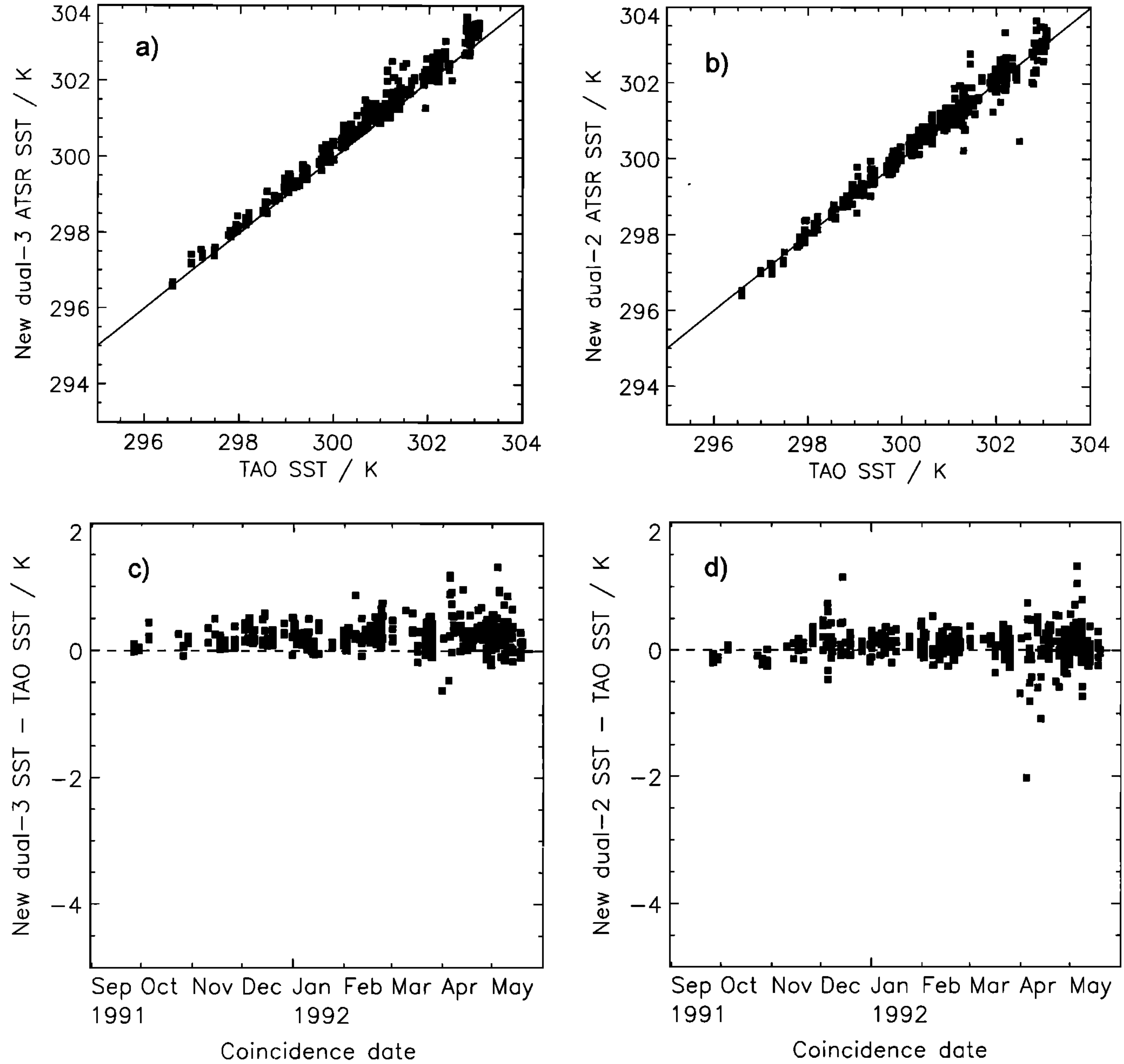


Figure 8. Comparisons of SSTs retrieved with new coefficients and in situ SSTs after bulk to skin adjustment. (a) New dual-3 retrievals against in situ skin SSTs. (b) New dual-2 retrievals against in situ skin SSTs. (c) Time series of differences between dual-3 retrievals and in situ skin SSTs. (d) Time series of differences between dual-2 retrievals and in situ skin SSTs.

The mean calculated skin-bulk temperature difference is 0.22 K (skin cooler than bulk), and this is reflected in the change in mean and median temperature differences. The systematic and random error in calculated skin-bulk differences are estimated to be ±0.04 K and ±0.07 K, respectively. Our best estimates of the biases in the new retrieval coefficients when applied to tropical regions are therefore ~0.2 K warm for dual-3 retrievals and ~0.07 K warm for dual-2 retrievals. Using conventional statistics, the root mean square of differences for the dual-3 and dual-2 retrievals is 0.22 and 0.27 K, respectively. The robust statistics indicate that for the majority of the coincidences (~95%), root-mean-square differences of better than 0.2 K are achieved by both retrieval algorithms.

What are the origins of the observed systematic and random errors (i.e., absolute bias and standard deviation) in the SST retrievals? First, we consider the biases, ~0.2 K warm for dual-3 retrievals and ~0.07 K warm for dual-2 retrievals. The standard error on these estimates of bias from the limited sample size is ~0.02 K; therefore the biases are significantly different from zero. Systematic error in the process of bulk-skin adjustment could be ~0.04 K. The possible errors in the humidity fields used in deriving the new coefficients cause negligible bias in dual-3 retrievals and a maximum bias of +0.04 K in dual-2 retrievals (see our companion paper). Simulations [*Merchant*, 1999] demonstrate that global dual-view coefficients (as used here) give a small cool bias at tropical and high latitudes, with a small warm bias at midlatitudes. The simulated cool (negative) bias in tropical regions is of magnitude ~0.07 K for dual-2 retrievals and <0.01 K for dual-3 retrievals. This effect thus explains about half of the observed difference in bias between the dual-2 and dual-3 retrievals but does not explain why both types of retrieval give results that are slightly warm. There has been speculation that the spectral response of the 3.7 μm channel shifted by -0.031 μm between calibration and operation [*Mason*, 1991], by some unknown mechanism. If true, this could introduce a warm bias in dual-3 retrievals of the order of 0.05 K, thereby accounting for the remainder of the observed difference in bias between the dual-2 and dual-3 retrievals. After allowing for these potential bias sources, an unaccounted for retrieval bias of at least 0.1 K remains. This suggests that some systematic error(s) yet remain(s) in the forward modeling process used to obtain the coefficients. Possibilities include remaining uncertainties in surface emissivity and spectroscopy, including the spectroscopy of tropospheric aerosols. However, it is worth emphasizing that all these residuals are small; in particular, achieving in an SST validation exercise a bias of less than 0.1 K with dual-2 coefficients under extreme aerosol and water vapor loadings represents a considerable success.

Second, we consider the factors contributing to the observed standard deviation between retrievals and the in situ skin SST estimate (Table 5). (1) Our estimate for random error in skin-bulk correction is 0.07 K (see above), and this adds to the error in bulk SST of 0.03 K. (2) The ATSR overpass is matched to the TAO measurements within an hour, and because we use only night data, the root mean square change in observed TAO SSTs over that timescale is small, ~0.03 K. (3) The effect of the difference in location and spatial scales of the retrieval (average over the clear-sky part of a 10 arc min cell with one corner at the TAO buoy) and the point in situ measurements cannot be directly assessed, there being no area-averaged in situ measurements available. However, the scatter between retrievals for adjacent 10 arc min cells matched to a single TAO measurement can be used to derive an upper bound for the error associated with variability in SST over spatial scales ~10 km. We find that 0.10 K is the average of the standard deviation between adjacent dual-3 SSTs, for the 60 cases where four ABTs are matched with a skin in situ SST. Assuming estimates (1) to (3) are all reasonable, the contribution to the observed standard deviation from factors other than intrinsic retrieval errors is ~0.13 K.

3 Summary and Conclusions

The prime purpose of this validation exercise is to test two aspects of the new coefficients for 10 arc min skin SST retrieval from ATSR, namely, their robustness to the influence of stratospheric aerosol and their performance under conditions of high water vapor loading. We use nighttime coincidences of ATSR overpasses with buoys in the tropical Pacific from the first 9 months of routine ATSR operation (before the 3.7 μm channel failure). The coincident data thus include the greatest stratospheric optical depth present during routine ATSR operation and total clear-sky water vapor ranging from 20 to 60 kg m^{-2}.

The usefulness of the 3.7 μm channel in SST retrieval from satellite is demonstrated by the good performance of all of the dual-view three-channel retrieval coefficients, giving low bias (magnitude 0.2 K or less) and low standard deviation (0.25 K), with no detectable trend from the changing stratospheric aerosol optical depth. Against the in situ measurements, the new retrieval coefficients were unbiased to a few hundredths of kelvin, but after adjustment of the in situ temperatures to skin SST, they appear to be biased warm by ~0.2 K, a result consistent with the findings of *Barton et al.* [1995] and *Mutlow et al.* [1994] for the prelaunch dual-3 retrieval scheme. This bias suggests that some systematic error yet remains in the forward modeling process used to obtain the coefficients. Small residual biases of this nature can readily be corrected empirically, but it is preferable to understand their origin and correct the systematic forward model error. For this tropical validation data set, dual-3 retrievals are about 0.15 K warmer than dual-2 retrievals. This result is consistent with the findings at tropical latitudes of the interalgorithm comparison conducted in our companion paper (see Figure 7 therein).

Our new retrieval coefficients bring a marked improvement in accuracy to dual-view two-channel retrievals of SST because of their aerosol-robust design and incorporation of the results of updated modeling of water vapor continuum absorption. A cold bias of the order of -1.5 K was present in dual-view two-channel retrievals using the prelaunch scheme during the period of this study. In contrast, by adjusting the in situ bulk SSTs to a skin temperature, we show that the new dual-2 retrievals are absolutely accurate to within 0.1 K. The standard deviation of retrieval-buoy differences using the new dual-2 coefficients is 0.27 K. This figure represents matchup error, buoy error, and skin-bulk adjustment error (with an estimated combined error of 0.13 K) in addition to the retrieval error intrinsic to use of the new coefficients. We note that the standard deviation is influenced by a small percentage of data which do not share the Gaussian distribution of the majority, suggesting the presence of uncleared cloud contamination in the averaged brightness temperatures. Calculation of a robust measure of standard deviation suggests that the retrieval coefficients give SSTs with standard deviation less than 0.2 K when unaffected by such errors. Within the statistical sensitivity of the exercise, there is no trend evident in the retrieved SSTs corresponding to the temporal trend in stratospheric aerosol optical depth. Together with the reduced retrieval biases and the results of the interalgorithm comparisons in our companion paper, this suggests that a high degree of aerosol-robustness has been achieved. In contrast, an aerosol-related trend is evident in the retrievals using the dual-2 Z95 (prelaunch) coefficients. The use of an updated parameterization of water vapor continuum absorption in deriving the new coefficients leads to retrievals that have no detectable sensitivity to high water vapor loadings over the range 20 to 60 kg m^{-2}, in contrast to retrievals using B97. Thus we regard the results presented here as a successful validation of the techniques used to derive the new coefficients described in our companion paper [*Merchant et al.*, this issue].

Acknowledgments. The authors thank the TAO Project Office, Michael J. McPhaden, director, for buoy data, and R. Grainger and A. Lambert of Oxford University for data from UARS. They also thank J. A. D. Mansley of the Mullard Space Science Laboratory for extracting microwave brightness temperature and water vapor data.

References

Analytical Methods Committee, Robust statistics—How not to reject outliers: 1 Basic Concepts, *Analyst, 114*, 1693-1697, 1989.

Barton, I. J., A. J. Prata, and P. Cechet, Validation of the ATSR in Australian waters, *J. Atmos. Oceanic Technol., 12*, 290-300, 1995.

Bernard, R., A. Le Cornec, L. Eymard, and L. Tabary, The microwave radiometer on board ERS-1, I, Characteristics and performances, *IEEE Trans. Geosci. Remote Sens., 31*, 1186-1198, 1993.

Brown, S. J., A. R. Harris, I. M. Mason, and A. M. Zavody, New aerosol-robust sea surface temperature algorithms for the along-track scanning radiometer, *J. Geophys. Res., 102*, 27,973-27,990, 1997.

Eymard, L., L. Tabary, E. Gérard, S.-A. Boukabara, and A. Le Cornec, The microwave radiometer aboard ERS-1, II, Validation of the geophysical products, *IEEE Trans. Geosci. Remote Sens., 34*, 291-303, 1996.

Fairall, C. W., E. F. Bradley, J. S. Godfrey, G. A. Wick, J. B. Edson, and G. S. Young, Cool-skin and warm-layer effects on sea surface temperature, *J. Geophys. Res., 101,* 1295-1308, 1996a.

Fairall, C. W., E. F. Bradley, D. P. Rogers, J. B. Edson, and G. S. Young, Bulk parameterization of air-sea fluxes for tropical ocean-global atmosphere coupled-ocean atmosphere response experiment, *J. Geophys. Res., 101,* 2747-2764, 1996b.

Freitag, H. P., Y. Feng, L. J. Mangum, M. J. McPhaden, J. Neander, and L. D. Stratton, Calibration procedures and instrument accuracy estimates of TAO temperature, relative humidity and radiation measurements, *NOAA/PMEL Tech. Rep. 1589,* Natl. Oceanic and Atmos. Admin., Silver Spring, Md., 1994.

Grant, A. L. M., and P. Hignett, Aircraft observations of the surface energy balance in TOGA-COARE, *Q. J. R. Meteorol. Soc., 124,* 101-122, 1998.

Harris, A. R., and M. A. Saunders, Global validation of the along-track scanning radiometer against drifting buoys, *J. Geophys. Res., 101,* 12,127-12,140, 1996.

Harris, A. R., M. A. Saunders, J. S. Foot, K. F. Smith, and C. T. Mutlow, Improved sea surface temperature measurements from space, *Geophys. Res. Lett.,22,* 2159-2162, 1995.

Hayes, S. P., L. J. Mangum, J. Picaut, A. Sumi, and K. Takeuchi, TOGA-TAO: A moored array for real-time measurements in the tropical Pacific Ocean, *Bull. Am. Meteorol. Soc., 72,* 339-347, 1991.

Lambert, A., R. G. Grainger, C. D. Rodgers, F. W. Taylor, J. L. Mergenthaler, J. B. Kumer, and S. T. Massie, Global evolution of the Mt. Pinatubo volcanic aerosols observed by the infrared limb-sounding instruments CLAES and ISAMS on the Upper Atmosphere Research Satellite, *J. Geophys. Res., 102,* 1495-1512, 1997.

Liu, W. T., K. B. Katsaros, and J. A. Businger, Bulk parameterisation of air-sea exchange of heat and water vapor including molecular constraints at the interface, *J. Atmos. Sci., 36,* 1722-1735, 1979.

Mason, G., Test and calibration of the along track scanning radiometer, a satellite-borne infrared radiometer designed to measure sea surface temperature, D. Phil. thesis, Univ. of Oxford, Oxford, England, 1991.

Merchant, C. J., Eliminating bias in satellite retrievals of sea surface temperature, Ph.D. thesis, Univ. of London, London, England, 1999.

Merchant, C. J., A.R. Harris, M. J. Murray, and A. M. Závody, Toward the elimination of bias in satellite retrievals of sea surface temperature, 1, Theory, modeling and interalgorithm comparison, *J. Geophys. Res.,* this issue.

Murray, M. J., M. R. Allen, C. T. Mutlow, A. M. Závody, M. S. Jones, and T. N. Forrester, Actual and potential information in dual-view radiometric observations of sea surface temperature from ATSR, *J. Geophys. Res., 103,* 8153-8165, 1998.

Mutlow, C. T., A. M. Závody, I. J. Barton, and D. T. Llewellyn-Jones, Sea surface temperature measurements by the along-track scanning radiometer on the ERS-1 satellite: Early results, *J. Geophys. Res., 99,* 22,575-22,588, 1994.

Saunders, P. M., The temperature difference across the cool skin of the ocean, *J. Atmos. Sci., 24,* 269-273, 1967.

Thomas, J. P., and J. Turner, Validation of Atlantic Ocean sea surface temperatures measured by the ERS-1 along track scanning radiometer, *J. Atmos. Oceanic Technol., 12,* 1303-1312, 1995.

Závody, A. M., C. T. Mutlow, and D. T. Llewellyn-Jones, A radiative transfer model for sea surface temperature retrieval for the along-track scanning radiometer, *J. Geophys. Res., 100,* 937-952, 1995.

A. R. Harris, U.K. Meteorological Office, London Road, Bracknell, Berks, RG12 2SZ, England. (arharris@meto.gov.uk)

C. J. Merchant, Department of Meteorology, The University of Edinburgh, James Clerk Maxwell Building, The King's Buildings, Mayfield Road, Edinburgh, EH9 3JZ, Scotland. (chris@met.ed.ac.uk)

(Received September 2, 1998; revised March 1, 1999;
accepted March 30, 1999.)